Maximum Likelihood Estimation of Multinomial Distribution

Setting :- $\mathcal{X} = \{x_1, x_2, \ldots, x_N\}$ N observations.

x_i :- a category out of K categories.

:- $[0, 0, 0, \ldots, \underset{\uparrow \text{kth index}}{1}, 0, \ldots, 0]$

:- One hot representation

- Assume an underlying Multinomial distribution.
- $\mathcal{X}$ was generated by sampling 'N' times from the multinomial.

Question :- What is the multinomial distribution that generated $\mathcal{X}$?

Some notation :-

Let $m = [m_1, m_2, \ldots, m_K]$ be a count vector on K categories.

$$m_1 + m_2 + \cdots + m_K = N$$

then a multinomial distribution is a '2' parameter distribution

$$P(m \mid \pi, N) = \text{Multinomial}(m \mid \pi, N)$$

$$= \binom{N}{m_1\, m_2 \ldots m_K} \prod_{k=1}^{K} \pi_k^{m_k}$$

$m = x_n = [0, 0, 0, \ldots, 1, 0, \ldots, 0]$

$N = 1$

and $\binom{1}{0\ 0 \ldots 1 \ldots 0} = 1$

So $P(x_n \mid \pi, 1) = \prod_{k=1}^{K} \pi_k^{x_{n,k}}$

Now writing $\mathcal{L}(\pi)$ for the dataset $\mathcal{X}$

$$\mathcal{L}(\pi) = \log \prod_{n=1}^{N} P(x_n \mid \pi, 1)$$

$$= \log \prod_{n=1}^{N} \prod_{k=1}^{K} \pi_k^{x_{n,k}}$$

$$\mathcal{L}(\pi) = \log \prod_{k=1}^{K} \pi_k^{\left(\sum_{n=1}^{N} x_{n,k}\right)}$$

$$= \sum_{k=1}^{K} \left(\sum_{n=1}^{N} x_{n,k}\right) \log \pi_k$$

maximize $\mathcal{L}(\pi)$ under the constraint $\sum \pi_k = 1$

Using lagrange multiplier λ to handle the constraint

So maximize

$$Obj(\pi) = \sum_{k=1}^{K} \left(\sum_{n=1}^{N} x_{n,k}\right) \log \pi_k + \lambda\left(1 - \sum_k \pi_k\right)$$

$$\frac{\partial}{\partial \pi_k} Obj(\pi) = \left(\sum_{n=1}^{N} x_{n,k}\right) \frac{1}{\pi_k} - \lambda = 0$$

$$\implies \pi_k = \frac{\sum_{n=1}^{N} x_{n,k}}{\lambda}$$

Solve for λ using

$$\sum_k \pi_k = 1$$

$$\frac{1}{\lambda} \sum_{k=1}^{K} \sum_{n=1}^{N} x_{n,k} = 1$$

$$\lambda = \frac{1}{\sum_{k=1}^{K} \sum_{n=1}^{N} x_{n,k}} = \frac{1}{N}$$

Hence $\quad \pi_k = \sum_{n=1}^{N} x_{n,k} \Big/ N$